Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A computer implemented method of identifying ~~and extracting~~ desired content ~~from~~ in HTML formatted web pages, comprising the steps of:

selecting a model page, wherein the model page includes content data and a plurality of HTML tags for formatting the content data;

identifying a first area of interest in the model page;

parsing the model page to generate a first string of symbols ~~corresponding to each of~~ for the plurality of HTML tags, the generated symbols in the first string representing only HTML tags, wherein the first area of interest is identified by a first portion of the first string of symbols;

retrieving a second web page associated with a different URL than the model page;

parsing the second web page to generate a second string of symbols ~~corresponding to each of the~~ for a plurality of HTML tags of the second web page, the generated symbols in the second string representing only HTML tags; and

comparing the first and second symbol strings to determine whether the second string includes a second portion similar to the first portion of the first string, wherein the second portion corresponds to a second area of interest in the second page.

2. (Original) The method of claim 1, wherein the step of comparing includes applying an approximate pattern matching algorithm to the first and second strings.

3. (Original) The method of claim 1, further comprising the step of storing the first and second areas of interest in a database.

4. (Currently amended) The method of claim 1, further comprising the step of extracting content data in the second area of interest from the second page.

5. (Original) The method of claim 4, further comprising the step of applying a regular expression matching algorithm to the extracted second area of interest.

6. (Original) The method of claim 1, wherein the first and second areas of interest each include two or more distinct sub-areas of the respective page.

7. (Original) The method of claim 1, wherein the step of identifying a first area of interest includes the step of identifying portions of the HTML tags of the model page.

8. (Original) The method of claim 1, wherein the step of identifying a first area of interest is performed using a manual pointing and selecting device.

9. (Original) The method of claim 1, wherein the steps of selecting and identifying are performed manually and wherein the remaining steps are performed automatically.

10. (Original) The method of claim 1, wherein the second web page is retrieved from a remote website over the Internet.

11. (Original) The method of claim 1, wherein the HTML tags include attributes and attribute values.

12. (Currently amended) A computer readable medium containing instructions for controlling a computer system to automatically identify ~~and extract~~ desired content ~~from~~ in a retrieved HTML formatted web page, by automatically:

parsing the HTML code of a manually selected model web page to generate a first string of symbols ~~corresponding to each of~~ for a first plurality of HTML tags, the generated symbols in the first string representing only HTML tags;

retrieving a second web page associated with a different URL than the model web page;

parsing the HTML code of the second web page to generate a second string of symbols ~~corresponding to each of the~~ for HTML tags of the second page, the generated symbols in the second string representing only HTML tags; and

comparing the first and second symbol strings to determine whether the second page includes a second plurality of HTML tags substantially matching the first plurality of HTML tags.

13. (Original) The computer readable medium of claim 12, wherein the first plurality of HTML tags are identified by an operator using a pointing and selection device coupled to the computer system.

14. (Original) The computer readable medium of claim 12, wherein the second web page is retrieved from a remote website over the Internet.

15. (Original) The computer readable medium of claim 12, further including instructions for extracting a portion of the second page corresponding to the second plurality of HTML tags.

16. (Original) The computer readable medium of claim 15, wherein the instructions further control the computer system to store the extracted portion of the second page in a database.

17. (Original) The computer readable medium of claim 15, further including instructions for controlling the computer system to apply a regular expression matching algorithm to the extracted portion of the second page.

18. (Original) The computer readable medium of claim 15, wherein the extracted portion of the second page includes two or more distinct sub-areas.

19. (Original) The computer readable medium of claim 12, wherein the instructions for comparing include instructions for applying an approximate string matching algorithm to the first and second strings.

20. (Original) The computer readable medium of claim 12, wherein the HTML tags include attributes and attribute values.

21. (Currently amended) A computer system for identifying and extracting content from HTML formatted web pages, the system comprising:

means for retrieving web pages including content data and HTML tags for formatting the content data, wherein a model web page is retrieved;

means for manually identifying a first area of interest in the model page, wherein the first area of interest corresponds to a first plurality of HTML tags; and

a processor including:

means for parsing a page, wherein the parsing means parses the model page and generates a first string of symbols ~~corresponding to each of~~ for the first plurality of HTML tags, the generated symbols in the first string representing only HTML tags, and wherein the parsing means thereafter parses an automatically retrieved second web page associated with a different URL than the model page and generates a second string of symbols ~~corresponding to each of the~~ for HTML tags of the second web page, the generated symbols in the second string representing only HTML tags;

means for comparing the first and second symbol strings to determine whether the second string includes a second portion similar to the first portion of the first string, wherein the second portion corresponds to a second area of interest in the second page; and

means for extracting content data in the second area of interest from the second page.

22. (Currently amended) A computer implemented method of identifying ~~and extracting~~ desired content ~~from~~ in web pages formatted using a markup language, comprising the steps of:

selecting a model page, wherein the model page includes a plurality of tokens, wherein tokens include HTML tag elements and content elements;

identifying a first area of interest in the model page;

parsing the model page to generate a first string of symbols ~~corresponding to each of~~ for the plurality of tokens in the model page, the generated symbols in the first string representing only tag elements, wherein the first area of interest is identified by a first portion of the first string of symbols;

retrieving a second web page associated with a different URL than the model page;

parsing the second web page to generate a second string of symbols ~~corresponding to each of the~~ for a plurality of tokens of the second web page, the generated symbols in the second string representing only tag elements; and

comparing the first and second symbol strings to determine whether the second string includes a second portion similar to the first portion of the first string, wherein the second portion corresponds to a second area of interest in the second page.

23. (Currently amended) The method of claim 22, further comprising the step of extracting content elements in the second area of interest from the second page.

24. (Original) The method of claim 22, wherein the markup language is selected from the group consisting of HTML, XML, WML, DHTML and HDML.

25. (Canceled).

26. (Currently amended) A computer-implemented method of identifying similar content in HTML formatted web pages, the method comprising:

selecting a model page, wherein the model page includes content data and a plurality of HTML tags for formatting the content data;

identifying a first area of interest in the model page;

generating a first string of symbols for the plurality of HTML tags associated with the first area of interest, the generated symbols in the first string representing only HTML tags; ~~each symbol corresponding to a different one of the plurality of HTML tags;~~

retrieving a second web page associated with a different URL than the model page;

generating a second string of symbols for the HTML tags of the second web page, the generated symbols in the second string representing only HTML tags; ~~each second symbol corresponding to a different one of the plurality of HTML tags of the second web page;~~ and

comparing the first and second symbol strings to determine whether the second string includes a portion similar to the first string, wherein the portion corresponds to a second area of interest in the second page.

27. (Currently amended) The method of claim 26, further comprising extracting content data in the second area of interest from the second page.

28. (Previously presented) The method of claim 26, wherein identifying is performed manually using a user-input device.